

## **Model Archive Summary for Suspended-Sediment Concentration at Station 11447903; Georgiana Slough Near Sacramento River, Ca**

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC beginning October 1, 2010. This is the first suspended-sediment model for this station. The methods used follow USGS guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.07/[2016.10](#) and USGS Techniques and Methods, [book 3, chap C4](#) (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (U.S. Geological Survey 2014).

### **Site and Model Information**

Site number: 11447903

Site name: Georgiana Slough near Sacramento River, California

Location: Latitude 38°14'14", longitude 121°31'03" referenced to North American Datum of 1983, Sacramento County, CA, Hydrologic Unit 18020109.

Equipment: The model covers the deployment of a YSI 6-series sonde equipped with a model 6136 turbidity sensor that began logging on December 3, 2009 until January 26, 2015.

Model number: 11447903.SSC.WY2011.1

Date model was created: September 13, 2013 and revised April 1, 2017

Model calibration data period: January 25, 2012 to December 14, 2014.

Model application date: October 1, 2010 to January 25, 2015.

Computed by: Tara Morgan-King USGS, Sacramento, CA [tamorgan@usgs.gov](mailto:tamorgan@usgs.gov)

Reviewed by: Anna Conlen USGS, Sacramento, CA [aconlen@usgs.gov](mailto:aconlen@usgs.gov)

### **Physical Sampling Details and Sediment Data**

All sediment data were collected using U.S. Geological Survey (USGS) protocols (U.S. Geological Survey, 2006) and are stored in the National Water Information System (NWIS) database <https://waterdata.usgs.gov/nwis>. Discrete, boat-based samples were collected seasonally, spanning the range of site conditions and specifically.

Discharge weighted suspended-sediment samples were collected along the transect, located roughly 200 ft downstream of the Georgiana Slough Bridge, using the Equal Discharge Method (EDI) to determine the locations of the sampling verticals. The EDI method was used to obtain the discharge-weighted sample because the site has rapidly changing tidal conditions and velocities and the samples are not always isokinetic due to the tidal nature of the site (based on Table 4-5 from [TWRI09A4, U.S. Geological 2006](#)). Each sampling vertical was located at the centroid of increments representing 20% of the total flow (5 verticals). A boat-based discharge measurement was collected immediately before sampling with an Acoustic Doppler Current Profiler (ADCP) to determine the location of each sampling vertical. A Federal Interagency Sedimentation Project (FISP) US D-96 bag sampler was used to collect depth-integrated samples

at each sampling vertical along the cross section by USGS personnel. The US D-96 bag sampler was equipped with a teflon nozzle. Any potential sampling bias due to non-isokinetic sampling (that can occur with the presence of suspended sand) is considered minimal (sand/fine break is described below). The channel cross section is approximately 30 feet deep in the thalweg with a mean sampling depth of approximately 27 feet. Velocities during the model calibration data period ranged from -0.3 to 3.1 ft/s

Samples were analyzed for SSC (mg/L) by the filtration method at the USGS Sediment Laboratory. Many samples were also analyzed for the percentage of fines (<0.063 mm). Sediment at this station is mostly fines (94% fines on average). The sand/fine analysis was useful to help identify potential sampling errors and/or outliers. The depth integrated samples collected from each of the 5 verticals were not composited for analysis but were instead analyzed individually primarily for quality control purposes as it enables analysis of each individual sample and associated errors that could otherwise skew the set average. Additionally, rapidly changing conditions in a tidal estuary could cause discrepancies in the sediment concentration between verticals so analysis at each vertical, helps to determine potential lateral variability. Once the SSC from each vertical is validated, the set average SSC from the 5 verticals in the cross section was computed to use in the calibration dataset. In rare occasions where the SSC at a vertical was deemed erroneous, a manual average is computed from fewer than 5 verticals and notes applied to the database.

On some sampling dates, multiple cross-sectional samples were collected. These samples are not considered replicates, as they were collected in a tidal system with rapidly changing conditions. The sediment lab auto-generates an event average for more than one set per day – thus two sets are automatically averaged on the same date. Generally, the data from event averages were not included in the calibration dataset if the collection times for the averaged samples were collected greater than 45 minutes apart. The event average was only used in the model dataset for the samples collected on 3/19/2012, as the two set averages were less than 45 minutes apart (0943 and 1018 Pacific Standard Time (PST)), and for the sample on 4/1/2012 for the same reason (set averages at 1013 and 1053 PST).

All sediment data were reviewed and marked as approved in the USGS NWIS Water-Quality System database (QWDATA) before being applied in the calibration dataset. The sediment results and metadata are stored and publicly available in the NWIS database located at <https://waterdata.usgs.gov/nwis> (U.S. Geological Survey, 2006).

## **Surrogate Data**

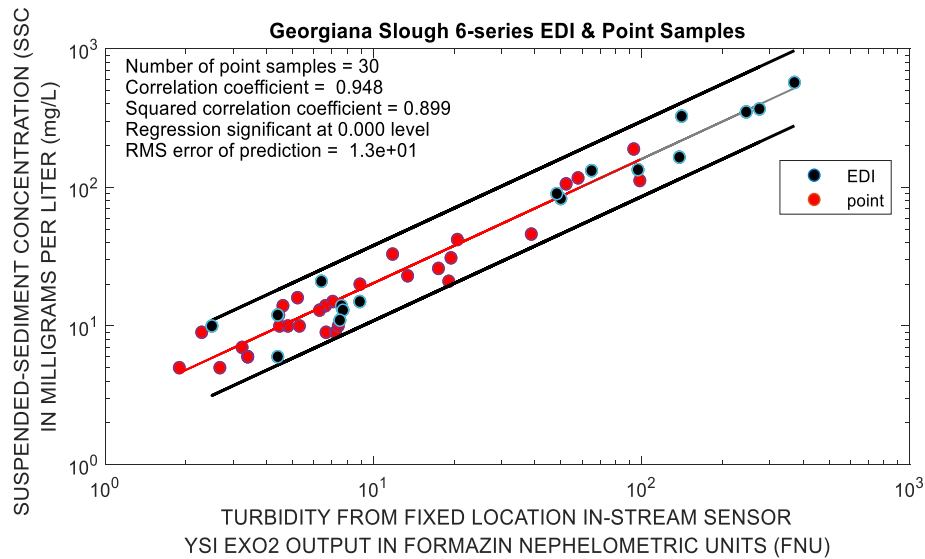
Continuous, 15-minute turbidity data, reported in Formazin Nephelometric Turbidity Units (FNU) and discharge data, reported in cubic feet per second (cfs), were evaluated as explanatory variables for SSC. Collection of turbidity data follow methods in Wagner and others (2006). Methods to compute discharge follow Levesque and Oberg (2012). Turbidity and discharge data were collected by the USGS California Water Science Center and are located at [https://waterdata.usgs.gov/nwis/uv?site\\_no=11447903](https://waterdata.usgs.gov/nwis/uv?site_no=11447903).

The approved time-series turbidity data spanning the dates of the discrete sediment sample dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were selected as a match for each sediment sample observation from a matching max +/- of 15 minutes for turbidity and discharge. An initial +/- 30 minute window was used to investigate models including tidally filtered discharge which only has hourly values in NWIS-TS. The SAID manual is found at <https://pubs.er.usgs.gov/publication/ofr20151177>.

### **Surrogate Turbidity with relation to Cross Section and Fixed-Point Samples:**

Some occasional point samples were collected over the years, but not in the traditional manner to compute a “box coefficient” that is used by the USGS in other programs. The box coefficient relationship is usually established by describing the ratio of the cross-section sample to the point sample (both sampled concurrently). Point samples were collected at the Georgiana Slough site at the location of the turbidity sensor but not concurrent with EDI cross-sectional samples. Three point-samples were collected during the 2010 water year, eight in water year 2011, sixteen in water year 2012, twenty-two in 2013, four in 2014, and five in 2015. Some of the samples are site visit pairs that were intended to describe potential changing conditions throughout a site visit. After multiple samples were rejected and corresponding turbidity values deleted for various reasons during records review, the final dataset consists 30 point-samples during the YSI 6-series deployment.

The relationship between turbidity to the point samples and the cross-sectional samples is similar (plot below). None of the point samples collected at the station were added to the calibration dataset, and were instead used to validate the model calibrated to cross-sectional, depth-integrated samples. The following graph demonstrates two things: 1) the point sample SSC compares with the cross-sectional SSC as the regression trend lines are nearly identical, and 2) even with fewer samples, the EDI cross-sectional calibration data set covers a higher range of turbidity conditions. Although the point samples cover more of the mid-range turbidity conditions, adding them does not improve the model.



## Calibration Dataset Development

The calibration data set includes sediment samples collected over 3 years from January 25, 2012 to December 14, 2014. We compiled a dataset of 18 observations. The number of turbidity and SSC pairs was reduced to 18 after one SSC sample was rejected in the database and two more observations were excluded due to a lack of corresponding turbidity values that were deleted during records computation. Data gaps in the turbidity record occurred for the samples on 8/22/2012 (16 mg/L) and 11/2/2012 (5 mg/L). Additionally, the EDI set average SSC on 3/19/2012 and on 4/1/2012 was used because these sets were collected less than 45 minutes of each other. As mentioned in the above section, the point samples were not combined with the cross-sectional samples just to reach the recommended sample size of  $n=36$ , however a test was done that determined it did not improve model error.

One sample was not included in the model computation from a composite EDI sample on 2/23/2012 @1016 (6 mg/L) because the SSC was low compared to the corresponding turbidity (in Formazin Nephelometric Turbidity Units, 13.4 FNU). The cross-sectional sample was determined erroneous based on the comparison with two point-samples collected the same day during equivalent turbidity conditions. The two point-samples had substantially higher SSC values of 23 and 24 mg/L at times of 1400 and 1515 PST, respectively. Turbidity values during these samples varied between 13.4 FNU and 13.8 FNU.

## Model Development

Simple linear regression (SLR) models and multiple linear regression (MLR) models were assessed using methods described in Helsel and Hirsh (2002). We ultimately tested roughly 20 different model variations, evaluating diagnostic statistics for differing linear and log models while aiming to retain as many of the observations as possible.

Model diagnostic statistical tests were performed using Matlab, SAID, and the R environment. Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. A script was created in Matlab to evaluate 5 models at once: Model 1) linear model

with one explanatory variable (turbidity), Model 2)  $\log_{10}$  transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5)  $\log_{10}$  model with two explanatory variables (turbidity and discharge). Because some observations were flagged in the model with 18 observations in SAID, we tested a variety of other models including only 17 observations, and only 16 observations. For each of the tests ( $n=18$ ,  $n=17$ , etc), each of the 5 models and diagnostic statistics were evaluated in Matlab. Note that most of the diagnostic statistics cannot be used to compare regressions with different response variable units, but in general, the highest  $R^2$ , lowest Mean Square Prediction Error (MSPE) values, and normality of residuals were considered optimal. The discharge surrogate was not significant as a second variable to the model ( $p>0.05$ ) and was not considered further.

Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to - 3 range are considered potential extreme outliers. The studentized residuals were reviewed from the SAID output reports and none of the samples were deemed as extreme outliers.

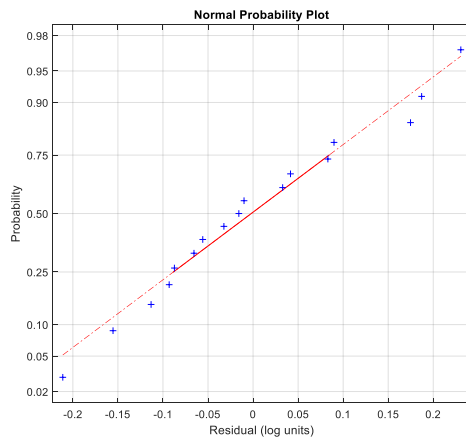
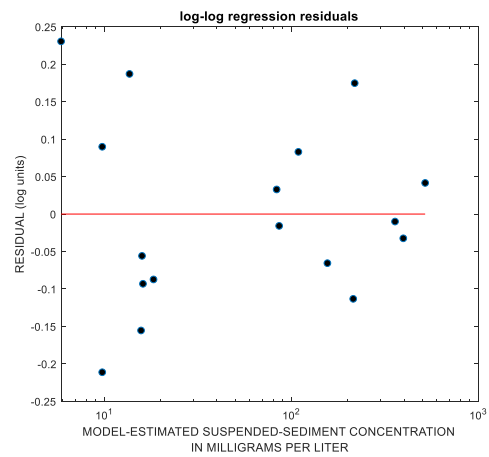
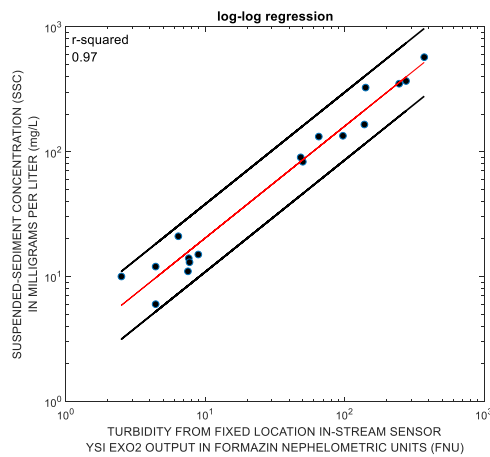
The diagnostic statistics improved with the decreased number of observations and were generally better for the linear models. The  $\log_{10}$  models had fewer flagged observations and better normality of residuals. The high sediment concentration samples were continually flagged in SAID. We evaluated the model by testing the removal samples that were flagged as potential outliers defined by the test criteria (leverage, Cook's D, and Dfitts; see Helsel and Hirsh (2002) for a description of these regression diagnostics). The observation on 12/14/2014 was continually flagged as high leverage but this sample represents the maximum SSC of the dataset (571 mg/L) and should be retained in the model. Because the  $\log_{10}$  model from the original dataset of  $n=18$  only flagged the observation from 12/5/2014, thus we tested removing this observation from the model. The model improved when the flagged sample from 12/5/2014 was removed and it was determined the corresponding turbidity was unexpectedly low for the higher SSC sample collected during the rising tide and might not have been representative. The MSPE initially looked better for the linear model (Model 1) compared to the  $\log_{10}$  model (Model 2) but the  $\log_{10}$  model had a slightly better r-squared (see table below) and better normality of residuals (see middle row of figure below). The re-transformed MSE, RMSE statistics for model 2 were computed and were slightly lower than for the linear model (Model 1). The results from the re-transformed predicted SSC are: RMSE = 34.6 and MSPE = 25.36.

No.	$R^2$	$R^2_a$	RMSE	PRESS	MSPE	n	(type)
Model 1	0.96	0.96	35.1	23328	25.68	17	linear
<b>Model 2</b>	<b>0.97</b>	<b>0.96</b>	<b>0.13</b>	<b>0.324</b>	<b>29.74</b>	<b>17</b>	<b>log</b>
Model 3	0.96	0.96	35.5	23887	26.02	17	repeated median
Model 4	0.97	0.96	32.7	47577	23.99	17	multi-linear
Model 5	0.97	0.96	0.13	0.47	30.49	17	multi-log

In summary, the best SLR model was determined to be a  $\log_{10}$  model with turbidity as the surrogate (from 17 observations) because it had the highest  $R^2$ , lowest overall error, and best normality of residuals.

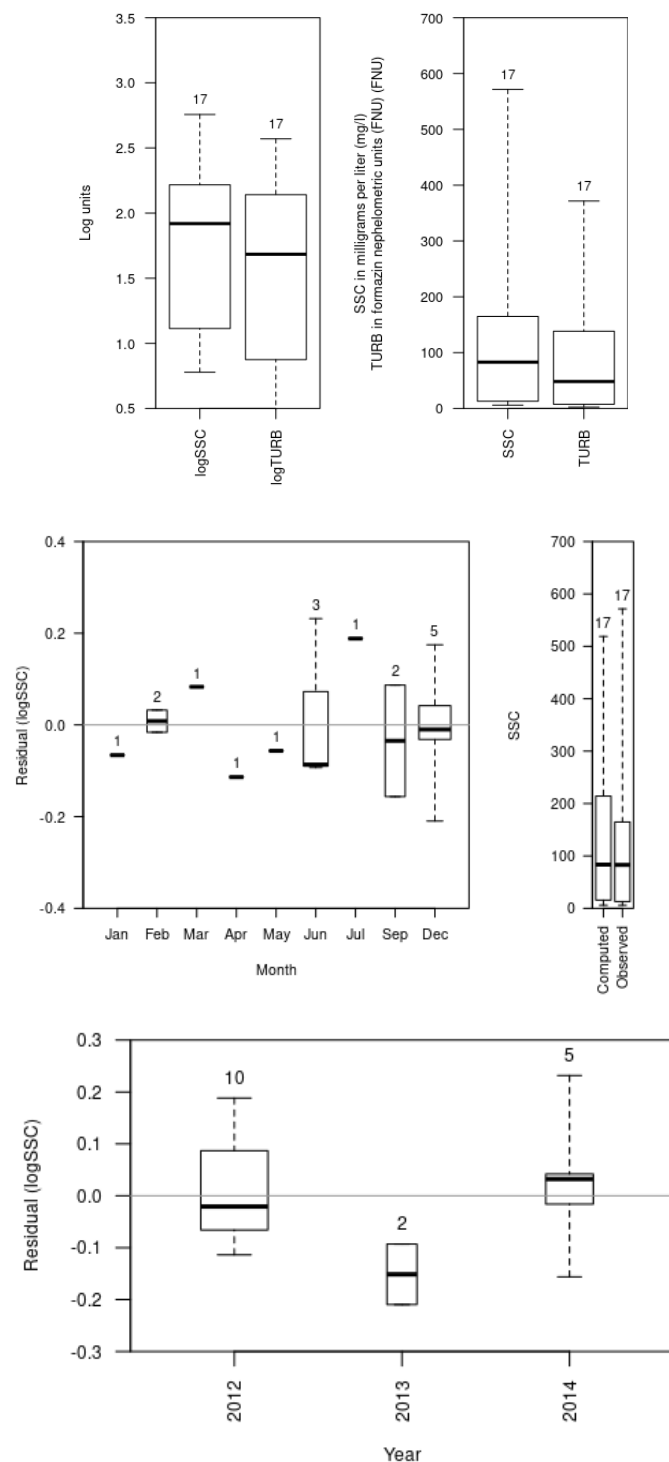
## Plots

The three plots below were output from Matlab and show the residuals for the log SLR model.



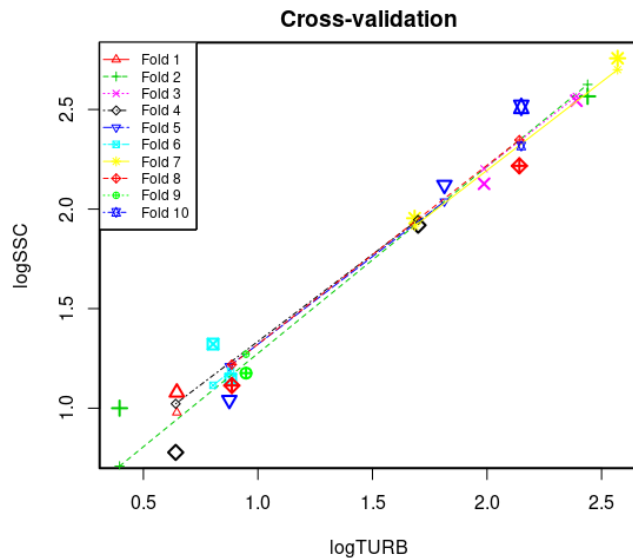
The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center, which is available at: <http://ksWSC.cr.usgs.gov:3838/peslick/ModelArchiveSummary/>.

Boxplots of turbidity and SSC data show the range of measured data for each model parameter. The second and third set of boxplots show SSC residuals by month by water year, respectively.

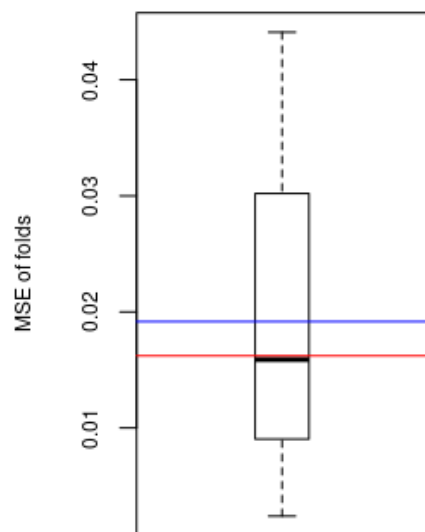


## Cross Validation

The cross-validation graph below shows a k-fold validation with k=10. The points represent observations that were left out of each fold. The bottom box plot show model MSE compared with the mean MSE of the folds.



Minimum MSE of folds: 0.0024  
Mean MSE of folds: 0.0192  
Median MSE of folds: 0.0159  
Maximum MSE of folds: 0.0441  
(Mean MSE of folds) / (Model MSE): 1.1800



Red line - Model MSE

Blue line - Mean MSE of folds



## Model Summary

The final SSC regression model is a  $\log_{10}$ -transformed SLR model based on 17 concurrent measurements of cross-sectional SSC samples and turbidity collected over approximately 3 years from January 25, 2012 to December 14, 2014. The model is shown below with basic model information, regression coefficients, and Duan's bias correction factor (Duan, 1973):

Linear Regression Model	Coefficient of Determination ( $R^2$ )
$\log_{10}SSC = 0.414 + 0.895 * \log_{10}Turb$	0.763

where

*SSC* = suspended-sediment concentration, in milligrams per liter (mg/L) and

*Turb* = turbidity, in formazin nephelometric units, measured with a YSI model 6136

SSC was transformed during regression model development, so the computed prediction may be biased. The  $\log_{10}$ -transformed SLR model can be retransformed and corrected for bias using the bias correction factor (BCF). The BCF for this model is 1.04.

Model	Start date	End date	Linear Regression Model	BCF
1	10/01/2010	01/25/2015	$SSC = 10^{0.414} \times Turb^{0.895} \times BCF$	1.04

or,

$$SSC = 2.70Turb^{0.895}$$

The SSC time-series is computed from USGS turbidity data. The computed time-series minimum and maximum values are shown below

Parameter	Minimum	Maximum
Computed SSC (mg/L)	1.7	550

\* The extrapolated, maximum allowable SSC for this model is 628 mg/L per guidelines. Extrapolation, defined as computation beyond the range of the model calibration dataset, may not be used to extend more than 10% outside the range of the sample data used to fit the model (U.S. Geological Survey, 2016). The maximum computed SSC from this model output time-series, was 550 mg/L thus, no data were removed from the record due to the extrapolation threshold.

## Suspended-Sediment Concentration Record

The SSC record is computed using the regression model on the USGS National Real-Time Water Quality (NRTWQ) website. The complete record can be found at: <https://nrtwq.usgs.gov/ca>.

### Model

$$\log\text{SSC} = + 0.895 * \log\text{TURB} + 0.414$$

### Variable Summary Statistics

Calibration data set (n=17)			Time-series data set	
SUMMARY STATISTIC	Turbidity, FNU	SSC, mg/L	Turbidity, FNU	SSC, mg/L
Minimum	2	6	0.6	1.7
1st Quartile	7	13	4	9.3
Median	48	83	6	14
Mean	87	137	13	25
3rd Quartile	139	205	6	21
Maximum	372	571	380	550

### Basic Model Statistics

#### Model Information:

Number of observations = **17**,  
root mean squared error (RMSE) = **0.13**,  
model standard percentage error (MSPE) = **29.74**,  
adjusted coefficient of determination ( $R^2_a$ ) = **0.96**,  
PRESS = **0.324**,  
bias correction factor = **1.04**

### Explanatory Variables

	Estimate	Standard error	t-statistic	p-value
Intercept	0.41404	0.070593	5.8652	3.11E-05
turbidity	0.89529	0.043252	20.699	1.92E-12

### Correlation Matrix

	Intercept	turbidity
Intercept	0.0049833	-0.0027455
turbidity	-0.0027455	0.0018707

### Outlier Test Criteria

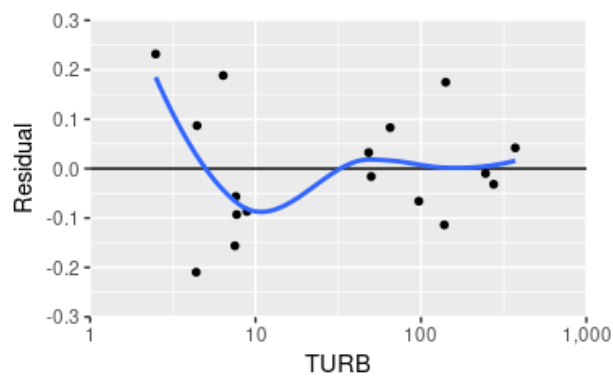
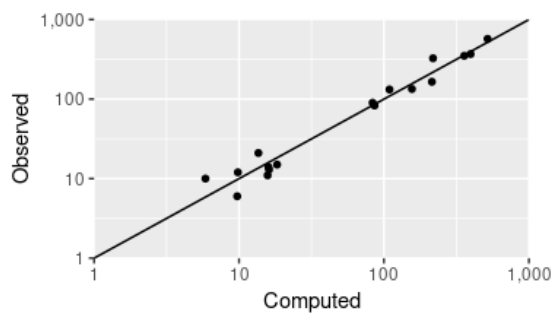
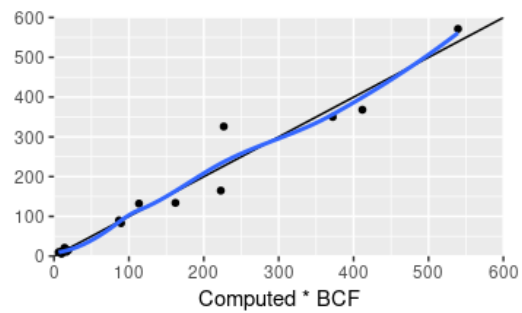
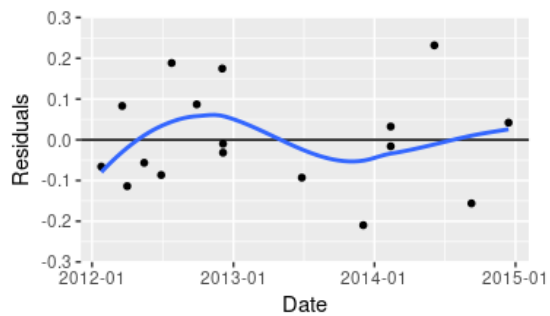
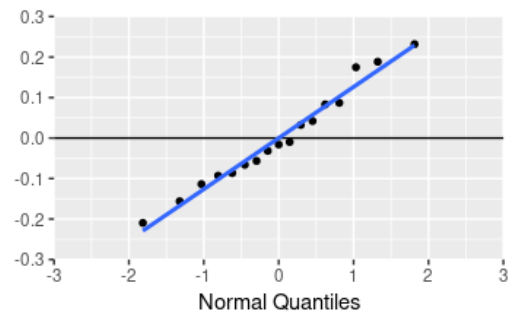
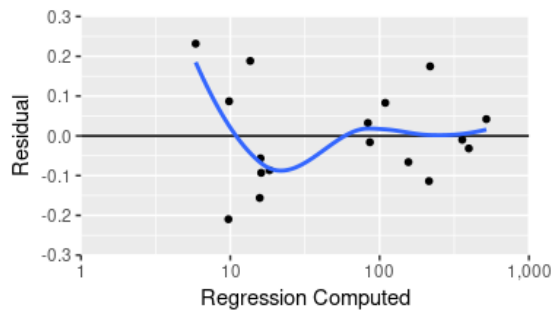
Leverage Cook's D DFFITS  
0.353 0.192 0.686

### Flagged Observations

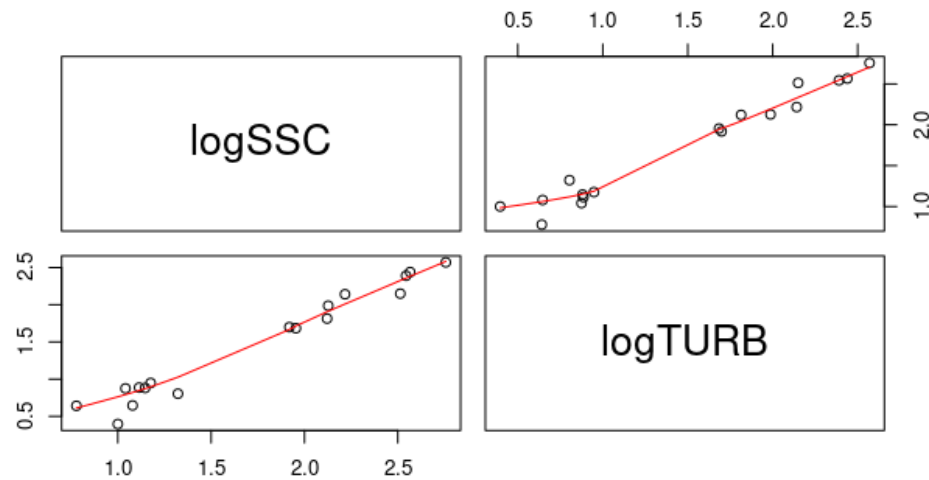
		logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's	D	DFFITS
12/3/2013	10:04	0.778	0.988	-0.210	-1.77	-1.93	0.138	0.251	-0.769	
6/5/2014	12:30	1.000	0.768	0.232	2.02	2.29	0.191	0.484	1.120	

## Statistical Plots

The following plots were generated per USGS model guidelines using the R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center. The app is available at: <http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/>



## Exploratory Plots



## Model Calibration Dataset

0	Date	logSSC	logTURB	SSC	TURB	Computed logSSC	Computed SSC	Residual	Normal Quantiles	Censored Values
1	2012-01-25	2.13	1.99	134	97.1	2.19	162	-0.0659	-0.452	--
2	2012-03-19	2.12	1.81	132	65.2	2.04	113	0.083	0.621	--
3	2012-04-01	2.22	2.14	165	138	2.33	223	-0.114	-1.03	--
4	2012-05-15	1.15	0.881	14	7.6	1.2	16.6	-0.0565	-0.296	--
5	2012-06-28	1.18	0.948	15	8.87	1.26	19	-0.0865	-0.621	--
6	2012-07-25	1.32	0.804	21	6.37	1.13	14.1	0.188	1.32	--
7	2012-09-28	1.08	0.646	12	4.43	0.992	10.2	0.0869	0.809	--
8	2012-12-03	2.51	2.15	326	141	2.34	227	0.175	1.03	--
9	2012-12-05	2.54	2.39	350	245	2.55	372	-0.00982	0.146	--
10	2012-12-05	2.57	2.44	368	275	2.6	412	-0.0317	-0.146	--
11	2013-06-27	1.11	0.886	13	7.69	1.21	16.7	-0.093	-0.809	--
12	2013-12-03	0.778	0.641	6	4.38	0.988	10.1	-0.21	-1.81	--
13	2014-02-12	1.92	1.7	83	50	1.94	89.5	-0.016	0	--
14	2014-02-12	1.95	1.68	89.9	48.3	1.92	86.8	0.0324	0.296	--
15	2014-06-05	1	0.396	10	2.49	0.768	6.1	0.232	1.81	--
16	2014-09-09	1.04	0.875	11	7.5	1.2	16.4	-0.156	-1.32	--
17	2014-12-14	2.76	2.57	571	372	2.72	539	0.042	0.452	--

## USGS Parameter Code Definitions

SSC: Suspended sediment concentration (SSC) in mg/l (80154)

TURB: Turbidity in FNU (63680)

USGS MAS App Version 1.0 available at:

<http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/>

## References

- Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., <http://dx.doi.org/10.3133/20151177>.
- Duan, Naihua, 1983, Smearing estimate-A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605-610.
- Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: [https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI\\_3-C2.pdf](https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf).
- Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.
- R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: <https://www.R-project.org/>.
- Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: <https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf>.
- U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: [https://pubs.usgs.gov/twri/twri9a4/twri9a4\\_Chap4\\_v2.pdf](https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf).
- U.S. Geological Survey, 2010, WRD Policy Numbered Memorandum No. 2010.02, Continuous Records Processing of all Water Time Series Data, Available from: <https://water.usgs.gov/admin/memo/policy/wrdpolicy10.02.html>
- U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: <https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf>
- U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: <https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf>
- Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data

reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available  
from: <https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf>.